

A joint source channel coding strategy for video transmission

C. PERRINE, C. CHATELLIER, S. WANG and C. OLIVIER

Laboratoire XLIM-SIC, Université de Poitiers
Téléport 2, Bvd M. et P. Curie, BP 30179, F-86962 Futuroscope-Chasseneuil cedex
Tel : +33 5 49 49 74 41
clency.perrine@sic.sp2mi.univ-poitiers.fr

Abstract – This paper presents a joint source channel coding designed for video transmission application. The aim is to improve the visual quality of the reconstructed video even when transmission errors occur while keeping the processing as simple as possible. The BER expected is higher than 10^{-4} .

This coding is based on an association of a wavelet transform (WT) and a vector quantization (VQ) optimally mapped on a QAM-M modulation. The whole of the transmission chain is jointly exploited in order to make it well-adapted to the complex and low data rate channel.

Keywords : Joint source channel coding, vector quantization, video transmission, GOP.

I. INTRODUCTION

Nowadays, most of telecommunication systems allow transmitting video in real time. Furthermore, the quality of the received video is an important parameter for the Quality of Service (QoS) supplied by the providers. This quality depends on the compression algorithm, the channel coding strategy and the transmission conditions in the wireless channel. The latest norms as H.264 or MPEG-4 [1] [2] have shown good performances related to the compression rate (optimized motion vector coding) and the scalability aspects. However, when errors occur over the wireless channel, video remains sensitive.

The aim of this paper is to present a joint source channel coding strategy based on a wavelet transform increasing the robustness of the transmission chain by reducing the impact of transmission errors (Bit Error Rate – $\text{BER} > 10^{-4}$) on the quality of restored images while contributing a good rate-distorsion. Wavelet-based coding is promising for efficient and scalable compression of video [7]. The real-time application constraint is taken into account insomuch as the complexity is reduced.

The video coding we suggest in this article is based on a vector quantization [3] for each sub-band of wavelet coefficients followed by an optimally QAM (16 or 256)

mapped strategy. This video coding is named WTSOM (Wavelet Transform Self Organized Map) video.

This algorithm based on fixed length coding has the advantage in opposition to the variable length coding (entropy coding) to be more robust regarding the transmission error but obviously with a lower compression rate.

The application expected in this study is videoconference which means limited scene change and low data rate channel.

In section II, the WTSOM video coding method is described, resorting to WTSOM strategy for fixed images [4] [10]. It is shown its adequacy with a digital QAM-256 modulation and the high performances related to the High BER. We propose, as for MPEG-4, to use the Group Of Pictures (GOP) layer containing 12 successive frames but from wavelet transform coefficients.

In section III, several transmission results through AWGN channels are presented involving the WTSOM video coding. The BER expected is higher than 10^{-4} . This video coding is more efficient in terms of robustness than in terms of compression aspects.

II. WTSOM VIDEO PRINCIPLE

After showing the difficulties to transmit video sequences when BER is high, the principle of WTSOM for fixed images is described [4] [10] and an extension to the video applications is detailed.

A. Issue

An example of video transmission is given in the figure 1. The standard of compression is MPEG4. It is transmitted through the AWGN channel using a QAM-16 modulation. The BER is closed to 10^{-4} . We can notice that

MPEG-4 encoding is extremely sensitive to errors due to the Variable Length Codes (VLC).

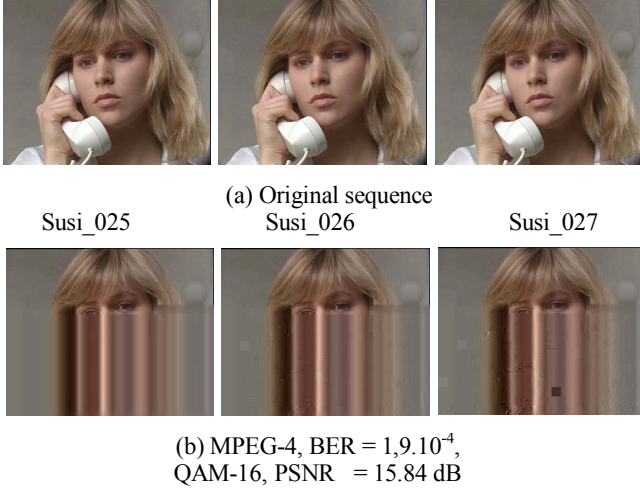


Figure 1. Transmission through ideal channel (without errors) and through AWGN channel:
(a) original sequence (b) MPEG-4, BER = $1,9 \cdot 10^{-4}$

In order to avoid this phenomenon, a source coding based on vector quantization is used which has the particularity to be a fixed length coding and well adapted to loss data compression. Before presenting the WTSOM algorithm for video, we suggest to give a brief description of WTSOM algorithm for fixed images [10].

B. WTSOM algorithm for fixed images

In our application, we use the Daubechies (9/7) biorthogonal wavelet such as in the JPEG 2000 case. The wavelet decomposition is applied at scale three and we preserve the five most significant sub images (LL_3 , HL_3 , LH_3 , HL_2 et LH_2) and discard the less important subbands. This choice is a good compromise between compression rate and visual quality of restored images. It is strengthened by an information entropy study for the 10 subbands related to our image sequence training basis. The size of the entire data before and after quantization is given in table 1.

The compression rate for this first step is 5,88 for any $n \times m$ image [11] :

$$T_c = \frac{n \times m}{(n/8 \times m/8) \times 3 + (n/8 \times m/8) \times 2} \approx 5,88 \quad (1)$$

The second stage is to apply a vector quantization on the remaining subbands of the wavelet decomposition. To do this, we use Kohonen's Self-Organizing Map (SOM) algorithm, because of its interesting property of auto-organization: close vectors in terms of Euclidian distance correspond to close indices in terms of location in the self-

organized codebook. This method is named WTSOM coding [5]. It requires 5 codebooks known by the transmitter and the receiver. The number and the vector shape depend on the subbands, the compression rate and the restored image quality expected. For our example in the figure 2, each codebook is composed of 256 vectors V_i for which the size is all the more important since the subband is more informative. The compression rate for the quantization step is 4,4 [11] with $n=m=512$ (see table 1) :

$$T_c = \frac{64 \times 64 \times 3 + 128 \times 128 \times 2}{64 \times 64 \times 2 + 128 \times 128 / 8} \approx 4,4 \quad (2)$$

Sub-band	Original size	Size after quantization
LL_3	$n/8 \times m/8$	$n/8 \times m/8$
HL_3	$n/8 \times m/8$	$(n/8 \times m/8)/2$
LH_3	$n/8 \times m/8$	$(n/8 \times m/8)/2$
HL_2	$n/4 \times m/4$	$(n/4 \times m/4)/16$
LH_2	$n/4 \times m/4$	$(n/4 \times m/4)/16$

Table 1 : Size of the entire data before and after quantization

The global compression rate is 25,6 ($5,88 \times 4,4$).

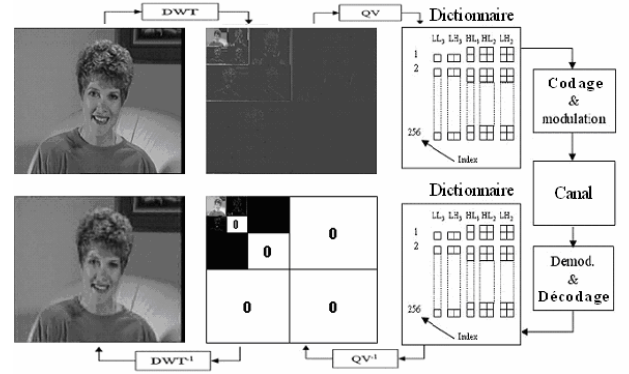


Figure 2. WTSOM video method :
VQ principle for 1 plan.

1) Robustness against transmission errors

Only the indexes of the codebook's vectors are transmitted through the channel using the right digital modulation. The size of the codebook is well-justified by the QAM-256 or QAM-16 modulation (figure 3). The codebook is optimally mapped on high spectral efficiency modulation and Kohonen topological maps (figure 3). They are more appropriate to QAM modulation because the layout of the codebook elements on the digital modulation points (in red) minimizes the average distortion during the transmission of codebook elements.

The robustness of the scheme suggested in this article is justified by the interdependence between the WTSOM video coding and the QAM modulation.

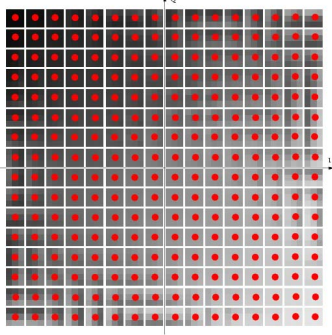


Figure 3. Codebook superimposing to QAM-256 modulation

In the figure 3, a 256-QAM has been used but in order to minimize the BER a 16-QAM modulation can be used. In this case, the optimal mapping of a 256-vector codebook is no longer possible. In order to solve this problem, Pyndiah [9] suggests reorganizing the codebook differently, that is to say using a four-dimension codebook (4x4x4x4) suitable for 4 bits (16-QAM). The results are not as good (1,5 dB loss in PSNR) as in the 256-QAM case but the system remains robust.

C. WTSOM video

In this section, WTSOM video (based on WTSOM for fixed images) is presented. This algorithm allows us to code image sequences by introducing GOP technique in order to increase the compression rate while keeping a good visual quality taking inspiration from the latest video standards. These techniques are applied to wavelet sub-band and a quantization.

1) Principle

We adopt the same strategy as they use in MPEG-4 standard, that is to say based on GOP technique using 12 successive images. Our method requires the construction of 2x5 codebooks obtained with differential video references : Mom-daughter, Claire, Missa, Grand-mom. For the codebooks, each vector has variable size as mentioned in section 2.2. The method we use based on GOP technique is performed from DWT sub-band.

In order to exploit the temporal redundancy GOP scheme is defined as follows:

- the I-frame (intra coded picture) reference picture corresponds to a fixed image and is independent of other picture types

- the P-frame (predictive coded picture) contains difference information from the preceding I or P-frame
- the B-frame (bidirectionally predictive coded picture) contains difference information from preceding and/or I or P-frame

The figure 4 illustrates the GOP structure with 13 frames.

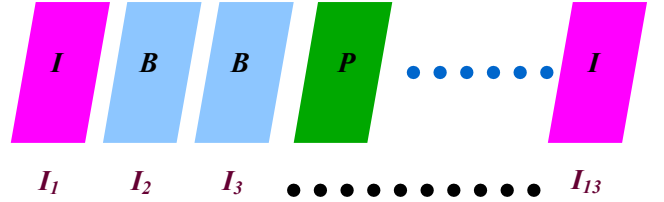


Figure 4. GOP structure with 13 frames

The reference pictures (I-frame) are transformed related to the DWT algorithm. These pictures are encoded according to the WTSOM algorithm for fixed images as mentioned before. These pictures will serve as reference to calculate P and B-frame and for decoding. The P-frames are predictively coded. A motion compensation of preceding pictures is done so that this image prediction could be realized. The B-frames are bidirectionally coded. Motion-compensated prediction is obtained with preceding and succeeding pictures.

Differential wavelet transform coefficients are calculated for each subband and between these images. $Dif_1, Dif_2, Dif_3, \dots$, will be the reference vectors. Thus, we are able to construct 5 codebooks for reference images and from difference vectors we are able to construct the new difference codebooks.

The difference equations are the following:

$$\begin{aligned}
 Dif_1 &= WTSOM(I_4) - WTSOM(I_1) \\
 Dif_2 &= WTSOM(I_2) - [WTSOM(I_1) + 1/3Dif_1] \\
 Dif_3 &= WTSOM(I_3) - [WTSOM(I_4) - 1/3Dif_1] \\
 Dif_4 &= WTSOM(I_7) - WTSOM(I_4) \\
 Dif_5 &= WTSOM(I_5) - [WTSOM(I_4) + 1/3Dif_1] \\
 Dif_6 &= WTSOM(I_6) - [WTSOM(I_7) - 1/3Dif_1] \\
 Dif_7 &= WTSOM(I_{10}) - WTSOM(I_7) \\
 Dif_8 &= WTSOM(I_8) - [WTSOM(I_7) + 1/3Dif_1] \\
 Dif_9 &= WTSOM(I_9) - [WTSOM(I_{10}) - 1/3Dif_1] \\
 Dif_{10} &= WTSOM(I_{11}) - [WTSOM(I_{10}) + 1/3Dif_1] \\
 Dif_{11} &= WTSOM(I_{12}) - [WTSOM(I_{13}) - 1/3Dif_1] \quad (3) \\
 &\dots
 \end{aligned}$$

where $WTSOM(I_j)$ means a wavelet coefficient of I_j plan. Dif_4, Dif_5, Dif_6 , etc ..., are obtained in the same way.

In the figure 5 is shown the WTSOM GOP principle.

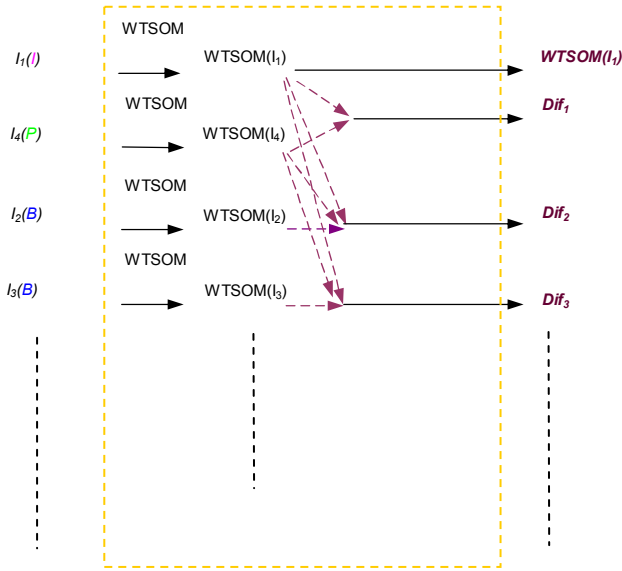


Figure 5. WTSOM GOP

I. VIDEO TRANSMISSION THROUGH AWGN CHANNEL

In this section, we show the results obtained with WTSOM video coding detailed above compared to MPEG-4 standard. The digital modulation is a QAM-16. Each symbol is coded using 4 bits what is easy to adapt to our method [9]. The main advantage, in comparison with QAM-256 modulation, is that it is easy to implement and more robust against the noise without disorganizing the initial scheme designed for 256 vectors (coded using $8 = 2 \times 4$ bits). The compression rate is about 44:1. The video "Mom" is transmitted over the AWGN channel (figure 6a : video clip from Mom: original : *Mom_024*, *Mom_025*, ... *Mom_035*). The codebooks are constructed in the same way as previously relatively to the chosen compression rate and from 4 basis video : *Mom-daughter*, *Claire*, *Missa*, *Grand-mom*.

The result of this sequence obtained with WTSOM video is given in the figure 6b. The restored sequence appears quite good in terms of visual quality. The PSNR equals 27, 28 dB , that is obviously less efficient than MPEG-4 or H264 standards.

The original sequence (figure 6a.) is then transmitted through an AWGN channel with the following features:

- $BER \approx 2.10^{-4}$
- $SNR = 8$ dB
- $T_c = 63.5$ for WTSOM video
- $T_c = 61$ for MPEG-4 standard
- Modulation : QAM-16

The results obtained with WTSOM video are illustrated in the figure 6c and those obtained with MPEG-4 in the figure 6d. In the figure 6c, the PSNR value remains constant compared to the figure 6b. In the figure 6e, we show a result with a high BER ($BER = 1,34.10^{-2}$) still using WTSOM video coding and with the same compression rate. This underlines the robustness of the method suggested in this article and its interest for video transmission through complex channels compared to MPEG-4 where the result is unacceptable.

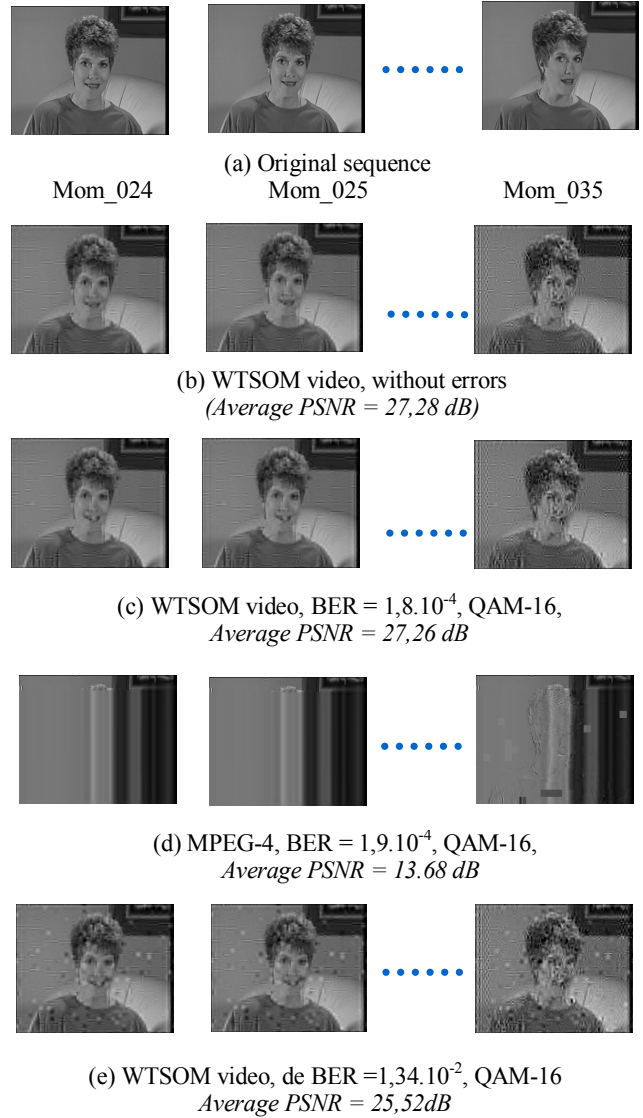


Figure 6. Sequence transmitted through ideal and AWGN channel comparable compression rate : (a) original sequence (b) WTSOM video, without errors, (c) WTSOM video, $BER = 1,8.10^{-4}$, (d) MPEG-4, $BER = 1,9.10^{-4}$ (e) WTSOM video, $BER = 1,32.10^{-2}$

IV. CONCLUSION AND OUTLOOKS

This paper presents a joint source channel coding designed for video application, the WTSOM video coding. It is based on an association of a wavelet transform, an appropriate codebook for each image of decomposition and a classical predictive coding involving a sequence of 12 images. We show that this strategy is particularly well-adapted to video transmission through complex and low data rate channel while keeping the useful bit rate as optimum. A comparison with MPEG-4 shows clearly the performances of WTSOM video in terms of visual quality and compression rate compromise.

The visual quality can be improved by introducing as for video standard motion vector what will allow a significant reduction in the residual distortion on the restored images [7]. Its encoding will be done by using vector quantization adapted to the modulation in order to keep the global strategy related to the robustness and the optimal useful bit rate.

Finally, some errors (figures 6c and 6e) could be attenuated by using video PDE (Partial Differential Equations)-based restoring algorithm as we developed [8] for fixed images.

REFERENCES

- [1] H264/MPEG-4 Part 10 Tutorials (Richardson), <http://www.vcodex.com/h264.html>
- [2] MPEG-4 Video Group, Coding of audio-visual objects: Video, ISO/IEC JTC1/SC29/WG11 N2202, March 1998
- [3] Aitsab O. "Turbo codes et codage conjoint source canal : application à la transmission d'images". PhD thesis, Ecole Nationale Supérieure des Télécommunications de Bretagne, Février 1998
- [4] Souhard, B. "Codage conjoint source canal : Application à la transmission d'images fixes sur canal ionosphérique", Thèse de l'Université de Poitiers, Mars 2004.
- [5] Kohonen T. "Self Organization and associative memory". Springer-Verlag, 1994.
- [6] Wang S., Chatellier C., et Olivier C. "Codage conjoint source canal pour des séquences d'images visioconférences", *CORESA*, Caen, pp 89-93, Novembre 2006
- [7] Agostini M.A., Antonini M., Barlaud M. "Model-based bit allocation between wavelet subbands and motion information in MCWT video coders", *EUSIPCO 2006*, Florence (Italy), September 2006
- [8] Bourdon P., Augereau B., Chatellier C., Olivier C.: "A multi-resolution, geometry-driven error concealment method for corrupted JPEG color images" - *Signal Processing: Image and Communication*, 20, pp 681-694, August 2005
- [9] J. Kangas, in: Increasing the error tolerance in transmission of vector quantize image by self organizing map, International Conference on Artificial Neural Networks, vol. 1, Paris, 1995, p. 287-291.
- [10] C. Chatellier, H. Boeglen, C. Perrine, C. Olivier, O. Haeberlé, "A robust joint source channel coding scheme for image transmission over the ionospheric channel" – *Signal Processing: Image Communication*, 22, pp 543-556, 2007.
- [11] S. Wang, "Stratégie de codage conjoint de séquences video base ondelettes" Ph.D thesis, University of Poitiers, 2008